Computational design of a homotrimeric metalloprotein with a trisbipyridyl core

Jeremy H. Mills[a,d,e,1], William Sheffler[a,1], Maraia E. Ener[a,b], Patrick J. Almhjell[d,e] Gustav Oberdorfer[a], José H. Pereira[f], Fabio Parmeggiani[a], Banumathi Sankaran[f], Petrus H. Zwart[f], and David Baker[a,c,2]

[a]Department of Biochemistry and the Institute for Protein Design, [b]Department of Chemistry, [c]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195; [d]School of Molecular Sciences, [e]The Biodesign Center for Molecular Design and Biomimetics, Arizona State University, Tempe, AZ, 85281 [f]Berkeley Center for Structural Biology, Physical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720.
[1]These authors contributed equally to this work
[2]To whom correspondence may be addressed:

David Baker
Molecular Engineering and Sciences Building
Box 351655
4000 15th Ave NE
University of Washington
Seattle, WA 98195
email: dabaker@uw.edu
Phone: 206-543-1295

Classifications: Biological Sciences, Biophysics and Computational Biology
Keywords: Computational protein design, Non-canonical amino acids, Metalloproteins

Abstract:

Metal chelating heteroaryl small molecules have found widespread use as building blocks for coordination-driven, self-assembling nanostructures. The metal chelating non-canonical amino acid ((2,2′-bipyridin-5yl)alanine (Bpy-Ala) could in principle be used to nucleate specific metalloprotein assemblies if introduced into proteins such that one assembly had much lower free energy than all alternatives. Here we describe the use of the Rosetta computational methodology to design a self-assembling homo-trimeric protein with $[Fe(Bpy\text{-}ala)_3]^{2+}$ complexes at the interface between monomers. X-ray crystallographic analysis of the homotrimer showed that the design process had near atomic level accuracy: the all atom RMSD between the design model and crystal structure for the residues at the protein interface is ~1.4 Å These results demonstrate that computational protein design together with genetically encoded non-canonical amino acids can be used to drive formation of precisely specified metal-mediated protein assemblies that could find use in a wide range of photo-physical applications.

Significance:

This article reports the computational design of a 3-fold symmetric, self-assembling protein homotrimer containing a highly stable non-canonical amino acid mediated metal complex within the protein interface. To achieve this result, recently developed protein-protein interface design methods were extended to include a metal chelating non-canonical amino acid containing a bipyridine functional group in the design process. Bipyridine metal complexes can give rise to photochemical properties that would be impossible to achieve with naturally occurring amino acids alone and suggests that the methods reported here could be used to generate novel photoactive proteins.

Introduction:

2,2'-bipyridine (Bpy) is one of the most widely used metal ligands in inorganic chemistry due to its redox stability, ability to form high affinity complexes of defined geometry with a variety of transition metals, and useful photochemical properties (1). The highly specific geometries adopted by Bpy in metal complexes has led to its widespread use in generating coordination-driven self-assembling nanostructures (2). For example, chemically synthesized, unstructured peptides containing the Bpy functional group as the side chain of an amino acid formed three helix bundles upon addition of metals (3, 4). Recently, the ability to use Bpy in biological contexts was expanded through the addition of the non-canonical amino acid (NCAA) (2,2'-bipyridin-5yl)alanine, (Bpy-ala) to the genome of *Escherichia coli* (5). Because Bpy forms very stable octahedral complexes with a number of biologically relevant divalent cations (e.g. Fe^{2+}, Zn^{2+}, Co^{2+}, and Ni^{2+}), an appropriately placed Bpy-ala residue could potentially be utilized to generate self-assembling proteins nucleated by $[M(Bpy)_3]^{2+}$ complexes (where M is a divalent cation that forms an octahedral complex with Bpy). Although we (6) and others (7, 8) have engineered proteins in which Bpy-ala served in structural or functional capacities, to our knowledge the possibility of using this NCAA to drive or stabilize the formation of a protein complex has not been explored.

The strategy of using metal ions to mediate protein complex formation has its origins in naturally occurring proteins such as hexameric proinsulin in which complex formation is regulated by bound Ca^{2+} and Zn^{2+} ions (9). Recently, the Tezcan (10) and Kuhlman (11) groups used rational and computational protein design methods, respectively, to engineer metal-dependent protein-protein interactions using conserved di-histidine metal binding motifs from known metalloproteins. In each case, structural analysis of the engineered protein complexes suggested that although metal dependent assembly had been achieved, the metal binding sites

had not formed as desired (10, 11). While notable successes, these studies highlight the difficulty in precisely stabilizing the conformations of multiple amino acid side chains at once. We reasoned that an appropriately placed, genetically encoded Bpy-ala residue could nucleate a 3-fold symmetric, homo-trimeric metalloprotein while overcoming some of the difficulties faced in the previous studies.

Here we report the use of the Rosetta computational protein design methodology (12) to engineer a homo-trimeric protein containing an octahedral $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex. Crystallographic analysis indicated that near atomic level accuracy between the design model and resulting protein was achieved in the vicinity of the protein-protein interface. The methods developed in this study could be applied to the generation of novel protein therapeutics, protein based materials, and metalloproteins with useful optical or photochemical properties.

Results

Computational design strategy

A previous study in which Bpy-ala was used to engineer a novel metal binding site in a protein scaffold highlighted important considerations for protein design efforts using this NCAA (6). In a design in which the Bpy-ala residue was not well constrained by packing interactions, we observed the formation of a highly stable $[Fe(Bpy\text{-}ala_f)_2(Bpy\text{-}ala_p)]^{2+}$ complex (where Bpy-ala_f is the free amino acid in the cell, and Bpy-ala_p is NCAA which has been incorporated in the protein) within the *E. coli* expression host. This observation was encouraging from the perspective that protein containing $[Fe(Bpy\text{-}ala)_3]^{2+}$ complexes can be purified intact from *E. coli*, but also suggested that for the present application, mono- or dimeric forms of the desired trimeric protein could be kinetically trapped by free Bpy-ala within the cell. Furthermore, even if a $[Fe(Bpy\text{-}ala_p)_3]^{2+}$ complex is formed as desired, in the absence of a driving force other than

the Bpy-ala complex formation, the subunit orientations in the resulting protein complexes would not likely be well defined.

To address these concerns, our computational design approach sought to achieve two distinct goals: identification of sites of Bpy-ala incorporation in starting scaffolds that are compatible with trimer formation, and design of interactions between protein subunits to drive the formation of a particular trimeric orientation at the exclusion of others. Our starting scaffold set consisted of a set of 9 *de novo* designed repeat proteins of known structure recently reported by our lab (13, 14). Because they are constructed from many identical subunits, the sizes of repeat proteins can be easily scaled in a defined manner and the versatility in length afforded by repeat protein architectures could be useful for downstream applications.

Potential sites of Bpy-ala incorporation within a given starting scaffold were first identified using Rosetta docking calculations (see supplemental information for full computational design methods). For each parent scaffold, 3 rotational and 1 translational degrees of freedom were sampled with a 2.5° covering radius to generate all possible 3-fold symmetric conformations in the context of polyalanine versions of the parent scaffolds. Successfully docked proteins contained geometrically complementary interfaces without steric clashes between the protein backbones of adjacent subunits.

We next searched for potential sites of octahedral $[M(Bpy\text{-}ala)_3]^{2+}$ complex incorporation within these trimeric scaffolds. Because previous results indicated that $[Fe(Bpy\text{-}ala)_3]^{2+}$ complexes can form within *E. coli* expression hosts (6), we chose to design our trimeric proteins in the context of these complexes. Octahedral Bpy complexes exist in two enantiomeric forms, Λ and Δ, and geometric constraints for each isomer were extracted from the Cambridge Structural Database (CSD ID KICRAR), and used to generate a $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex with

ideal geometries for further calculations (Fig. 1A). All symmetry related Cβ atoms in a given docked configuration were then computationally scanned for the ability to overlap with the Cβs in Λ- and Δ-Bpy-ala complexes. Placed Bpy-ala complexes were removed from further consideration if the Cα-Cβ-Cγ angle differed by more than 3° from the value of 116° observed in the crystal structure of a Bpy-ala containing protein (PDB ID 4iww).

Docked trimer models containing a $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex (Fig. 1B) were then subjected to iterative rounds of RosettaDesign (15) to generate a low energy, well-packed interface between protein subunits (Fig. 1C). These designs were then filtered on the shape complementarity (16) between protein interfaces and the change in solvent exposed surface area (ΔSASA) upon dissociation of the protein complex.

Initial Characterization of the Designed Proteins

A total of seven designed proteins were generated through oligonucleotide mutagenesis of the parent scaffolds. An amber stop codon (TAG) was substituted for the wild type codon at the desired site of Bpy-ala incorporation. pET21b expression plasmids encoding the designed proteins were co-transformed with the pEVOL-BpyRS plasmid (Schultz laboratory, The Scripps Research Institute) into a BL21 (DE3) *E. coli* expression strain. The pEVOL-BpyRS plasmid contains an orthogonal tRNA / aminoacyl tRNA synthetase pair that enzymatically acylates a tRNA containing an anti-codon loop specific to the amber codon with the Bpy-ala amino acid. Thus, full-length protein expression should only occur in the presence of the NCAA; in the absence of Bpy-ala in the expression medium the amber stop codon should terminate translation. Expression trials for each of the designed proteins were therefore carried out in the presence and absence of Bpy-ala. Of the 7 designs tested, 6 showed full-length protein expression only in the presence of the Bpy-ala amino acid (Fig. S1), and were subjected to additional characterization.

Tris-Bpy complexes containing metals including Zn^{2+}, Ni^{2+}, or Fe^{2+} have characteristic spectroscopic signatures in the UV and visible wavelength ranges. Absorbance in the range of 290-330 nm corresponds to $\pi—\pi^*$ transitions within the Bpy ligand, and is observed in complexes of Bpy with a number of biologically relevant metals (17). $[Fe(Bpy)_3]^{2+}$ complexes exhibit metal-ligand charge transfer (MLCT) which gives rise to characteristic absorption spectra in the visible wavelength range of 450-575 nm (18). To examine whether or not Bpy-ala mediated complex formation had occurred, the absorbance of the designs was analyzed from 230-600 nm after nickel affinity purification. Two designed proteins (TRI_03 and TRI_05) showed absorbance in the range of 450-575 nm with λ_{max} values of 490 and 525 nm, consistent with a $[Fe(Bpy)_3]^{2+}$ complex (Fig. 2A and 2C) (18). Both TRI_03 and TRI_05 had as their parent scaffolds *de novo* designed ankyrin like repeat proteins (PDB IDs 4gpm and 4hb5 respectively) (13). Purified designs were then subjected to analysis by size exclusion chromatography (SEC), and were observed to elute at volumes indicative of proteins slightly larger than expected for the trimeric complexes relative to standards of known molecular weight (Fig. 2B and 2D). This discrepancy in apparent size could be due to the unusually large Stokes radii of the trimeric complexes (19). Although both proteins formed some soluble aggregates (Fig. 2B and 2D), TRI_03 appeared to be more prone to aggregation than TRI_05. Chiral Λ and Δ Bpy complexes give rise to characteristic circular dichroism (CD) signals in the near UV. CD analysis of TRI_05 gave a spectrum characteristic of the Λ isomer (Fig. 2E) (18), which was consistent with the designed complex.

To further characterize TRI_03 and TRI_05, the proteins were expressed and purified in large scale in preparation for crystallographic analysis. TRI_03 was observed to precipitate at

high protein concentration, likely due to its propensity to aggregate. In contrast, TRI_05 was stable at high concentrations, and was subjected to crystallographic trials.

Structural Characterization of TRI_05

We solved the structure of TRI_05 to 2.2 Å resolution. The designed trimeric topology of TRI_05 was observed in the crystal structure and electron density corresponding to a tris-bipyridine metal complex was clearly visible at the trimeric interface (Figs. 3A and 3B). The RMSD between the design model and the crystal structure is 1.4 Å for all atoms on the first two helices that participate directly in the interface. When superimposed on the interface residues, global deviations between the design model and structure are observed in the C-terminal repeats, likely due to slight deviations in the vicinity of the designed interface that propagate into much larger differences in the remainder of the protein. A global fit of all backbone atoms in the design to the structure gives an RMSD of 2.5 Å. Very good agreement between the design model and the structure was observed for the side chains in the interface with the exception of Ile14, Leu38, and Met46. Ile14 adopts different rotamers in the design model and the solved structure (Fig. 3C). Larger deviations are observed for Leu38 and Met46, due to an unanticipated interaction between Met46 and Leu26 that forces Leu38 into a rotameric state not present in the design model.

Transient Absorbance Analysis of TRI_05

Femtosecond time resolved spectroscopic methods can be used to probe the photophysical properties of complexes such as $[Fe(Bpy)_3]^{2+}$ (20). Photoexcitation of such complexes in the MLCT domain results in the transfer of an electron from the metal ion to the ligands and the solvent environment can affect the evolution of the excited state (21). Because TRI_05 represents the first example of a protein containing a $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex, we

explored the possibility that the excited state dynamics would be altered by encapsulation of the complex within the protein environment. We therefore carried out femtosecond time resolved absorption spectroscopy on both TRI_05 and the free $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex. Excitation at 440 nm results in a characteristic loss of absorbance (bleaching) in the MLCT range (450-700 nm, Fig. 4A). Recovery of this bleach was used to determine excited state decay. Overlays of the excited state decays measured at 530 nm (corresponding to the wavelength with largest ΔA) of TRI_05 and the free $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex are essentially superimposable (Fig. 4B). To more fully quantify any differences, mono-exponential fits of these data were carried out, and suggested lifetimes of 690 ps for TRI_05 and 610 ps for the free complex (Fig. S2). These lifetimes are identical within error, and are in relatively good agreement with previously reported values of the excited state lifetime of $[Fe(Bpy)_3]^{2+}$ complexes (810 ps, ref 19); the small difference between our values and the previously reported lifetimes could originate from differences in Bpy substitution or solvent composition. One potential explanation for the similarity of our measured values for the excited state lifetimes of free $[Fe(Bpy\text{-}ala)_3]^{2+}$ and TRI_05 is that the $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex is only partially buried within the protein interface, and is solvent accessible from at least one face (Fig. 3 A and B). Additionally, the solvent-sensitive excited-state lifetimes of complexes such as $[Ru(Bpy)_3]^{2+}$ are orders of magnitude longer than the $[Fe(Bpy)_3]^{2+}$ complexes (in the range of 10-600 ns, ref. (20)). In contrast, the picosecond lifetimes of $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex in TRI_05 may simply be too short to report on the surrounding environment.

Metal Dependence of the Complex

TRI_05 was purified from the expression host in complex with Fe^{2+}. In order to analyze the dependence of the protein complex the bound metal, we attempted to remove the metal from

the protein by adding an excess of the potent chelator 1,10-phenanthroline (Phen), which has an affinity for Fe^{2+} ~4 orders of magnitude higher than Bpy (22). Incubation of TRI_05 with an excess of Phen at 25 °C did not result in rapid exchange of the two metal ligands. We reasoned that metal removal might be accelerated at increased temperatures, and thus examined the thermal stability of TRI_05 using CD analysis (Fig. S3A-D). Although slight unfolding was observed with increasing temperatures (Fig. S3B), TRI_05 did not appear to fully denature up to 95 °C (Fig. S3C). Furthermore, the partial unfolding was found to be reversible upon cooling to room temperature (Fig. S3C). TRI_05 was incubated with an excess of Phen for 1 hour at 65 °C, after which spectroscopic analysis indicated the presence of a $[Fe(Phen)_3]^{2+}$ complex, with no discernable signal indicative of $[Fe(Bpy)_3]^{2+}$ complex (Fig. S4 A and B). TRI_05 was then again subjected to SEC to remove unbound Phen and the $[Fe(Phen)_3]^{2+}$ complex. Apo TRI_05 eluted from the SEC column at the same volume observed prior to metal removal (Fig. S4 C), suggesting the presence of Fe^{2+} was not required to maintain the trimeric complex. Eluted TRI_05 was analyzed spectroscopically at high concentration and no signal in the range of 450-575 nm was observed suggesting the bound Fe^{2+} had been removed (Fig. S4 D). The possibility existed that apo TRI_05 scavenged metal, most likely Zn^{2+} during the purification process. To examine this, we carried out inductively coupled plasma mass spectrometric (ICP-MS) analysis of purified TRI_05 before and after removal of Fe^{2+} with Phen (Table S1). Although very little Zn^{2+} was observed in the Fe^{2+} containing trimer, a 2:1 ratio of Zn^{2+} to Fe^{2+} was observed in the apo protein suggesting the protein may have bound to Zn^{2+} during the purification process. Interestingly, no absorbance was observed in the range of 290—330 nm in the apo protein (Fig. S4 D) which would have been indicative of Zn^{2+} bound to Bpy-ala.

To further confirm that a bound metal ion was not required for complex formation, we co-transformed the plasmid encoding the TRI_05 gene with a plasmid containing a synthetase that suppresses the amber stop codon with tyrosine rather than Bipy-ala. This tyrosine containing protein was subjected to SEC analysis and was observed to elute at the same time as the metal containing TRI_05 suggesting the presence of bound metal is not required to maintain trimer formation (Fig. S4 E). Finally, we explored the possibility that apo TRI_05 associated in a concentration dependent manner. When apo TRI_05 was subjected to SEC at 750 µM, the protein eluted at a volume consistent with the previously observed trimer with a shoulder indicative of a higher order species. When diluted to 75 µM and 35 µM, apo TRI_05 eluted at volumes 1.4 and 2 mL later than the concentrated sample respectively (Fig. S4 F). This suggests that, in the absence of a bound metal ion, TRI_05 is capable of self-associating, but does so in a concentration dependent fashion.

We then examined the ability to re-form the $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex through addition of Fe^{2+} to solutions of apo TRI_05. Ferrous ammonium sulfate was added to a final concentration of 20 µM to a solution of apo TRI_05 at 60 µM and incubated for 12 hours at 65 °C for 1 hour. Spectroscopic analysis of the TRI_05 suggested that no re-metallation had occurred under these conditions. We hypothesized that this was due to formation of a metal free complex in which one or more Bipy-ala side chains were buried in the protein interface. We then diluted apo TRI_05 to a final concentration of 2.5 µM (below the concentration observed to elute as a monomer on SEC) and concentrated the protein in the presence of 10 µM ferrous ammonium sulfate via ultrafiltration. The concentrated protein had an absorbance spectrum consistent with the original TRI_05 (Fig. S4 G), and analysis by SEC suggested that the complex

had partially reformed whereas apo TRI_05 subjected to the same conditions in the absence of Fe^{2+} remained a monomer (Fig. S4 H).

Discussion

From the perspective of design of functional metalloproteins, the use of the Bpy functional group in a biological context provides a number of advantages. The Bpy ligand provides two pyridyl nitrogens in an orientation compatible with metal binding, and therefore has a higher inherent metal affinity than any naturally occurring amino acid. Bpy is also neutral and non-polar and hence it should be compatible with protein interfaces comprised mostly of hydrophobic residues. The energies of formation of octahedral complexes of Bpy with biologically relevant metals range from -18.0 kcal / mol for Zn^{2+} to -27.5 kcal / mol for Ni^{2+} (values calculated from affinities reported in (22)) rivaling those of the tightest known protein-protein interactions (23). However, one potential drawback to the use of Bpy-ala in this manner is that these energetically favorable complexes form very rapidly and do not readily exchange at neutral pH (24). Thus, while genetically encoded Bpy-ala could potentially be used to nucleate a C-3 symmetric, heterotrimeric protein complex around a bound metal, the potential to kinetically trap the protein complex as a monomer or dimer with free Bpy-ala, or form complexes with unintended orientations must be avoided.

We addressed these issues by using the Rosetta computational protein design methodology to identify sites of Bpy-ala incorporation in protein scaffolds of known structure that were sterically compatible with complex formation and to design novel protein-protein interfaces that provide an additional driving force for the formation of the desired complexes at the exclusion of other, undesired states. Of the seven computationally designed proteins we characterized experimentally, two were observed to have spectroscopic signatures indicative of

the formation of $[Fe(Bpy\text{-}ala)_3]^{2+}$ complexes. Determination of the structure of one of these designs, TRI_05, revealed a trimeric metalloprotein containing the desired $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex; high similarity between the solved structure and the design model was observed in the vicinity of the trimeric interface. Removal of the bound Fe^{2+} ion from TRI_05 with the potent metal chelator 1,10-phenanthroline did not result in the dissolution of the trimeric protein complex into its component monomers (Fig. S4 C). This result highlights the ability of computational protein interface design algorithms to engineer high affinity interfaces between proteins that can be used to drive the formation of complexes of desired orientations.

The computational methods developed in the course of this study could be extended to generate other novel proteins with useful properties: The use of other metals or metal chelating NCAAs (e.g. (8-hydroxyquino-lin-3yl)alanine (25)) could result in the formation of 2-fold symmetric protein-protein complexes with similar stabilities. The ability to site specifically place metal ions within the context of protein complexes could pave the way to the development of new biomaterials with useful properties. Finally, we demonstrated the ability to remove the and re-introduce the bound iron suggesting the possibility of introducing photoactive metals such as Ru^{2+} or Os^{2+} in the future, thereby facilitating the development of proteins with photochemical properties not achievable using exclusively naturally occurring amino acids.

Materials and Methods

Generation and cloning of designed proteins

Sequences of the designed proteins were backtranslated and optimized for expression in *E. coli* using DNAWorks (26). DNAWorks also generates sets of overlapping oligonucleotides from which full-length genes can be generated through assembly PCR (26). These oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA) and

assembled through a two-step PCR reaction. PCR assembled gene fragments were then cloned into a pET21 expression vector (Novogen) between NdeI and XhoI restriction sites. Sequences of the full-length genes were confirmed by Sanger sequencing.

Protein expression and purification

Slightly different protocols were used for analytical and production scale protein expressions. Detailed descriptions of each of these methods can be found in the SI; a general protein expression and purification protocol follows. Sequence confirmed pET21 expression plasmids were co-transformed with the pEVOL-BpyRS plasmid into chemically competent BL21(DE3) cells (Life Technologies) and were selected on LB-agar plates containing ampicillin and chloramphenicol antibiotics. Colonies with resistance to both antibiotics were used to inoculate expression cultures in Terrific Broth (TB) supplemented with both antibiotics. Expression cultures were grown at 37 °C until an optical density at 600 nm (OD_{600}) value of ~0.75 was reached. The temperature was dropped to 18 °C and the Bpy-ala NCAA was added to a final concentration of 150 μM. Expression was induced through the addition of L-arabinose and isopropyl β-D-1-thiogalactopyranoside to final concentrations of 0.02% and 1 mM respectively and growth was allowed to continue at 18 °C for 30 hours at which time cells were harvested via centrifugation.

Cell pellets were lysed via sonic dismembration and the lysates were clarified by centrifugation. The designed proteins were then purified by immobilized metal ion affinity chromatography on Ni-NTA resin (Qiagen) followed by SEC analysis on a Superdex 75 column (G.E. Biosciences). The SEC step served to further purify the protein and also indicated the size of the expressed proteins.

Crystallization of TRI_05

TRI_05 was dialyzed against 25 mM Tris buffer pH 8.0 containing 150 mM NaCl, and concentrated via centrifugal ultrafiltration to a final concentration of 12 mg / ml. Concentrated TRI_05 was screened using the sparse matrix method (27) with a Phoenix Robot and the following crystallization screens: Crystal Screen, SaltRx, PEG/Ion, Index and PEGRx (Hampton Research, Aliso Viejo, CA) and Berkeley Screen (Lawrence Berkeley National Laboratory). Crystals of TRI_05 were found in the Berkeley Screen condition consisting of 0.1 M Magnesium Chloride, 0.1 M Sodium Acetate trihydrate pH 4.5 and 30% Pentaerythritol propoxylate. TRI_05 crystals were obtained after 5 days by the sitting-drop vapor-diffusion method with the drops consisting of a mixture of 0.2 µl of protein solution and 0.2 µl of reservoir solution.

Metal removal from TRI_05

TRI_05 at concentrations between 50 and 75 µM, and in buffer containing 50 mM Tris, pH 7.6 and 150 mM NaCl was directly mixed with 1,10-phenanthroline to give a final chelator concentration of 1 mM. The protein and 1,10-phenanthroline mixture was then incubated at 65 °C for one hour. To separate TRI_05 from free 1,10-phenanthroline and the $[Fe(Phen)_3]^{2+}$ complex generated during the incubation, the mixture was subjected to gel filtration immediately after the incubation was complete.

Re-addition of Fe^{2+} to TRI_05

Apo TRI_05 was diluted to a concentration of 2.5 µM in buffer containing 50 mM Tris, pH 7.6 and 150 mM NaCl from which contaminating metals had been removed by treatment with Chelex 100 resin. Ferrous ammonium sulfate was added to the protein solution to a final concentration of 10 µM and the mixture was applied to a Vivaspin 20 centrifugal concentrator (Sartorius). The protein was concentrated via centrifugation at 4,500 x g to a final volume of 0.5 mL and was immediately subjected to analysis by SEC.

ICP-MS analysis of proteins

To prepare samples for ICP-MS analysis, TRI_05 bound to Fe^{2+} and apo TRI_05 were dialyzed extensively against 1.5 L of 50 mM Tris, pH 7.6, and 150 mM NaCl. In total, the buffer was changed 5 times. Prior to dialysis, beakers were soaked overnight in 2 mM EDTA in an effort to remove contaminating metals. Buffers were also treated with Chelex 100 (Biorad) resin to remove exogenous metals. Samples were digested for three hours in trace metal grade HNO_3 and H_2O_2 with heating. After drying, samples were further digested overnight in concentrated HNO_3 and HCl, and again were dried. Dilute HNO_3 was added to a final volume of 15 mL the samples were applied to the instrument. Analysis was carried out on an iCAP Q quadrupole Electron X-series ICP-MS instrument (Thermo) in kinetic energy discrimination mode.

X-ray diffraction collection and structural solution of TRI_05

X-ray diffraction data were obtained at the Advanced Light Source on beamlines 8.21 at wavelength 1.0 Å, and the initial data were processed with HKL2000 (28). Phase information was obtained with molecular replacement using PHASER (29); a single chain of the TRI_05 design in which all side chains were truncated at Cβ was used as a search model. Side chains were modeled using AutoBuild (30) using the sequence of the designed protein in which Bpy-ala was mutated to alanine. Bpy-ala was substituted for alanine in the AutoBuild generated model using PyMol (31). The model containing Bpy-ala was used as an input for a round of Rosetta-Phenix refinement (32), after which iterative rounds of structural refinement (including addition of water molecules) were carried out with PHENIX (33).

Transient Absorption Studies

For transient absorption (TA) studies, 50 femtosecond laser pulses were generated by a Libra Ti:Sapphire laser system (Coherent). Approximately 75% (3 W) of the Ti:Sapphire output was used to pump an OPerA Solo Optical Parametric Amplifier (Coherent) to generate 440 nm

excitation pulses, while the remainder (1 W) was reserved for visible probe generation. The excitation and probe beams were directed into a Helios TA spectrometer (Ultrafast Systems), where broadband, visible probe light was generated at a sapphire plate. TA spectra were collected in random order at 300 log-spaced time delays over the interval of ~2 ps to 5 ns; spectra at each time point were averaged over three scans. Data were processed with Surface Explorer software (Ultrafast Systems), and plotted and fit to single exponential decays using Matlab 2014b curve fitting software (MathWorks). Spectra of TRI_05 and $[Fe(Bpy\text{-}ala)_3]^{2+}$ were collected at concentrations of 100 µM in 30 mM Tris (pH 8.0) and 150 mM NaCl in quartz cuvettes. An attempt was made to degas the TRI_05 sample, but resulted in precipitation of the protein. TRI_05 spectra were therefore collected in cuvettes open to air. Degasing of the $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex was possible, and the effect of oxygen on the recorded spectra was examined. A $[Fe(Bpy\text{-}ala)_3]^{2+}$ sample was degased prior to analysis via repeated rounds of evacuation of a stoppered cuvette followed by backfilling with N_2. Data collected in the presence and absence of air were indistinguishable from one another (Fig. S5).

Acknowledgements

The authors would like to thank Peter Schultz for the generous gift of the pEVOL-BpyRS plasmid. We thank Neil P. King for helpful discussions, Prof. Cody Schlenker (Office of Naval Research DURIP grant number N00014-14-1-0757) for access to the ultrafast TA laser system, and Tim Pollock for experimental assistance. We thank Gwyneth Gordon and Trevor Martin for assistance with ICP-MS analysis. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under

Contract No. DE-AC02-05CH11231. J.H.M. acknowledges support from the National Institute of General Medical Science of the National Institutes of Health under award F32GM099210. D.B. and J.H.M. were supported by the Defense Threat Reduction agency under award HDTRA1-11-1-0041. M.E.E. acknowledges support from the ACS Irving S. Sigal Postdoctoral Fellowship. F.P. was the recipient of a Swiss National Science Foundation Postdoc Fellowship (PBZHP3-125470) and a Human Frontier Science Program Long-Term Fellowship (LT000070/2009-L). G.O. is a Marie Curie International Outgoing Fellowship fellow (332094 ASR-CompEnzDes FP7-People-2012-IOF)

Author Contributions: J.H.M., W.S., M.E.E., and D.B. designed research; J.H.M., W.S., M.E.E., J.H.P, F.P., B.S., and P.H.Z. performed research, J.H.M., M.E.E., G.O., B.S., P.H.Z., and D.B. analyzed data; J.H.M. and D.B. wrote the paper with input from all co-authors.

Data deposition: The atomic coordinates of TRI_05 have been deposited in the Protein Databank (www.pdb.org) under PDB ID 5eil.

The authors declare no conflicts of interest

References:

1. Kaes C, Katz A, Hosseini MW (2000) Bipyridine: The Most Widely Used Ligand. A Review of Molecules Comprising at Least Two 2,2'-Bipyridine Units. *Chem Rev* 100(10):3553–3590.
2. Leininger S, Olenyuk B, Stang PJ (2000) Self-Assembly of Discrete Cyclic Nanostructures Mediated by Transition Metals. *Chem Rev* 100(3):853–908.
3. Lieberman M, Sasaki T (1991) Iron(II) organizes a synthetic peptide into three-helix bundles. *J Am Chem Soc* 113(4):1470–1471.
4. Ghadiri MR, Soares C, Choi C (1992) A convergent approach to protein design. Metal ion-assisted spontaneous self-assembly of a polypeptide into a triple-helix bundle protein. *J Am Chem Soc* 114(3):825–831.

5. Xie J, Liu W, Schultz PG (2007) A Genetically Encoded Bidentate, Metal-Binding Amino Acid. *Angew Chem Int Ed* 46(48):9239–9242.

6. Mills JH, et al. (2013) Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *J Am Chem Soc* 135(36):13393–13399.

7. Lee HS, Schultz PG (2008) Biosynthesis of a Site-Specific DNA Cleaving Protein. *J Am Chem Soc* 130(40):13194–13195.

8. Kang M, et al. (2014) Evolution of iron(II)-finger peptides by using a bipyridyl amino acid. *Chembiochem* 15(6):822–825.

9. Dunn MF (2005) Zinc-ligand interactions modulate assembly and stability of the insulin hexamer -- a review. *Biometals* 18(4):295–303.

10. Salgado EN, Faraone-Mennella J, Tezcan FA (2007) Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J Am Chem Soc* 129(44):13374–13375.

11. Der BS, et al. (2012) Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J Am Chem Soc* 134(1):375–385.

12. Leaver-Fay A, et al. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth Enzymol* 487:545–574.

13. Parmeggiani F, et al. (2015) A general computational approach for repeat protein design. *J Mol Biol* 427(2):563–575.

14. Park K, Shen BW, Parmeggiani F, Huang PS (2015) Control of repeat-protein curvature by computational protein design. *Nat Struct Mol Biol* 22(2):167-174.

15. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA* 97(19):10383–10388.

16. Lawrence MC, Colman PM (1993) Shape Complementarity at Protein/Protein Interfaces. *J Mol Biol* 234(4):946–950.

17. Meyer TJ (1986) Photochemistry of metal coordination complexes: metal to ligand charge transfer excited states. *Pure Appl Chem* 58(9):1193-1206.

18. Mason SF (1968) The electronic spectra and optical activity of phenanthroline and dipyridyl metal complexes. *Inorg Chim Acta Rev* 2:89–109.

19. Hong P, Koza S, Bouvier ESP (2012) Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol* 35(20):2923–2950.

20. Creutz C, Chou M, Netzel TL, Okumura M, Sutin N (1980) Lifetimes, spectra, and

quenching of the excited states of polypyridine complexes of iron(II), ruthenium(II), and osmium(II). *J Am Chem Soc* 102(4):1309–1319.

21. McCusker JK (2003) Femtosecond absorption spectroscopy of transition metal charge-transfer complexes. *Acc Chem Res* 36(12):876–887.

22. Martell AE, Smith RM (1982) *Critical Stability Constants* (Springer US, Boston, MA) doi:10.1007/978-1-4615-6761-5.

23. Baxendale JH, George P (1950) The kinetics of formation and dissociation of the ferrous tris-dipyridyl ion. *Trans Faraday Soc* 46(0):736–9.

24. Horton N, Lewis M (1992) Calculation of the free energy of association for protein complexes. *Protein Sci* 1(1):169–181.

25. Lee HS, Spraggon G, Schultz PG, Wang F (2009) Genetic Incorporation of a Metal-Ion Chelating Amino Acid into Proteins as a Biophysical Probe. *J Am Chem Soc* 131(7):2481–2483.

26. Hoover DM, Lubkowski J (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* 30(10):e43–e43.

27. Jancarik J, Kim SH (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Crystallogr* 24(4):409-411.

28. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.

29. McCoy AJ, Grosse-Kunstleve RW (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(4):658-674.

30. Terwilliger TC, et al. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(1):61–69.

31. Schrödinger, LLC The PyMOL Molecular Graphics System, Version 1.7.2.2.

32. DiMaio F, et al. (2013) Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat Meth* 10(11):1102–1104.

33. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(2):213–221.

Figures and Legends:

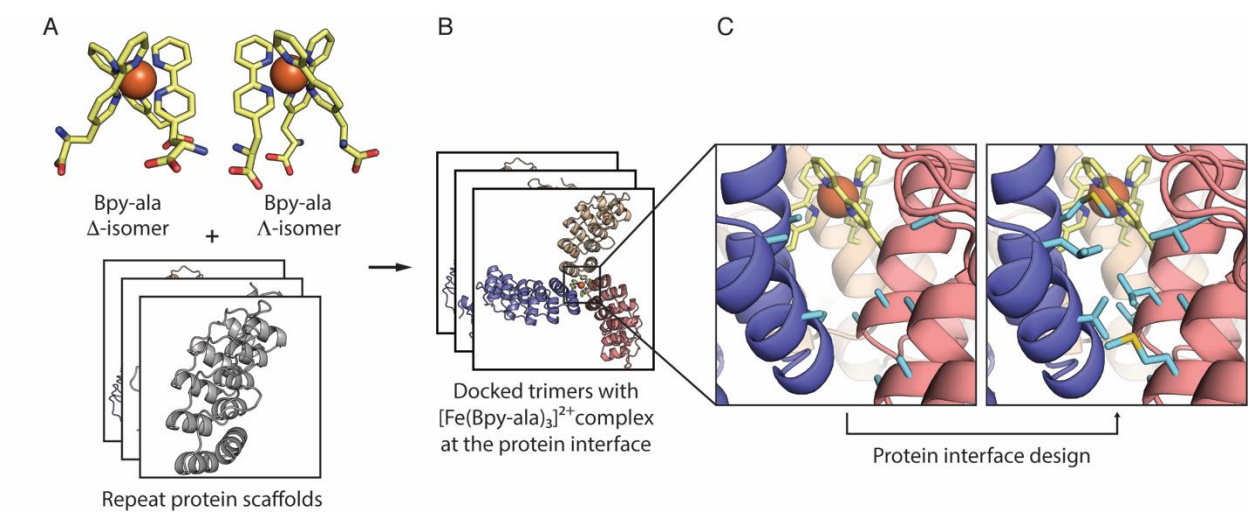

Fig. 1. An overview of the computational design methods. (A) Octahedral $[Fe(Bpy\text{-}ala)_3]^{2+}$ complexes (yellow sticks, Λ and Δ isomers) were generated from small molecule crystal structures and used as inputs for a docking algorithm that identified sites of incorporation in

repeat protein scaffolds (gray cartoons) compatible with a 3-fold symmetric protein complexes. (B) Successfully docked trimers (multi-colored cartoons) were compatible with the geometries set by $[Fe(Bpy\text{-}ala)_3]^{2+}$ complexes, and contained no steric clashes between protein subunits. (C) Computational interface design methods were used to engineer highly complementary surfaces between trimer subunits (light blue sticks) to drive the formation of a protein complex with a desired 3-fold symmetric orientation.

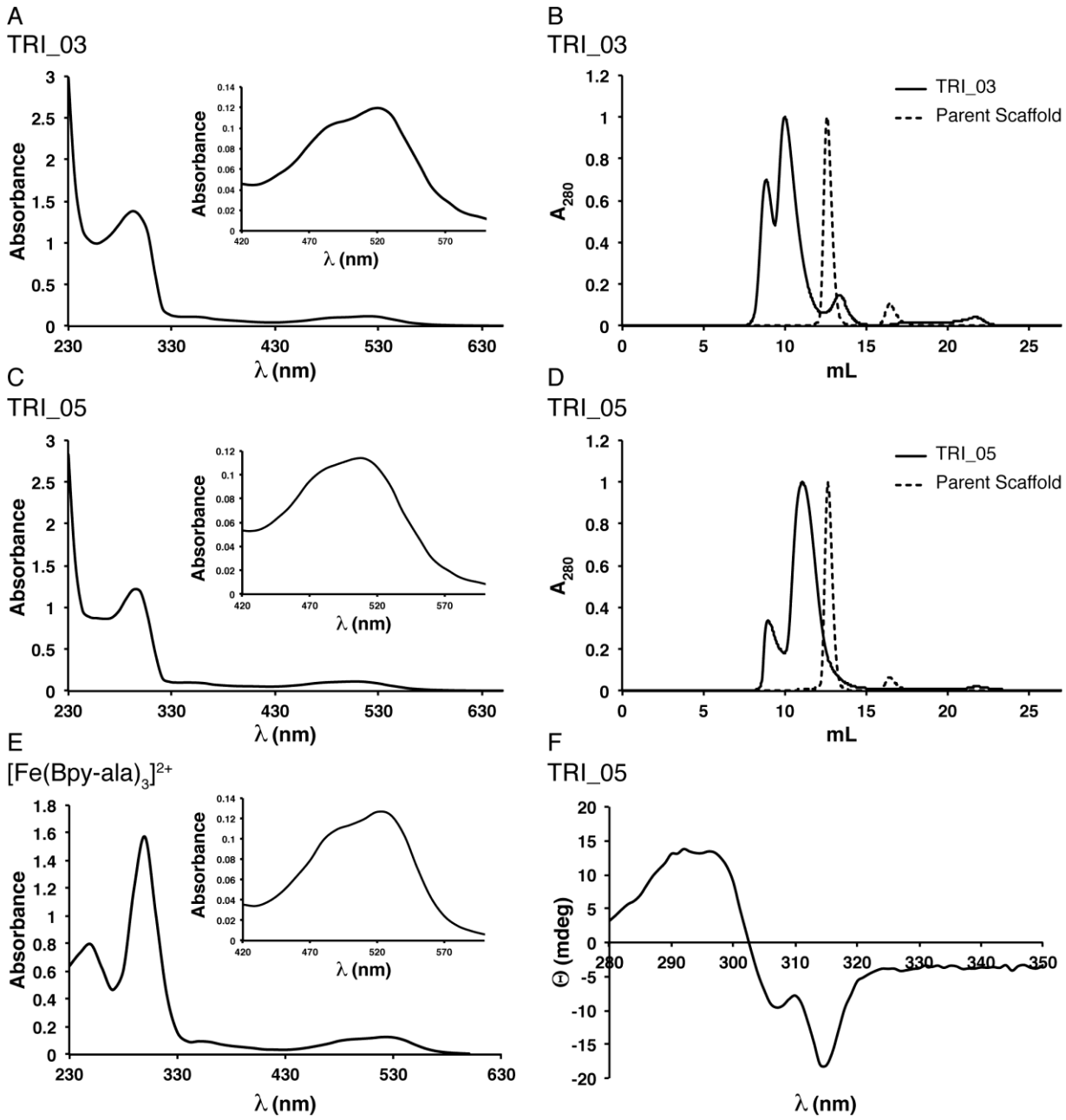

Fig. 2. Experimental characterization of designs TRI_03 and TRI_05. Spectroscopic characterization of TRI_03 (A) and TRI_05 (C) and $[Fe(Bpy\text{-}ala)_3]^{2+}$ (E) in the range of 230—650 nm is shown. The MLCT regions of these spectra are enlarged in the insets. SEC chromatograms of TRI_03 (B) and TRI_05 (D) are shown (solid lines) overlaid with traces of the monomeric parent scaffolds of TRI_03 and TRI_05 (PDB IDs 4gpm and 4hb5 respectively, dashed lines). Soluble aggregates are observed for each protein at the column void volume of ~8

mL. Near UV circular dichroism analysis of TRI_05 (F) gave a spectrum consistent with the Λ□isomer of a $[Fe(Bpy)_3]^{2+}$ complex.

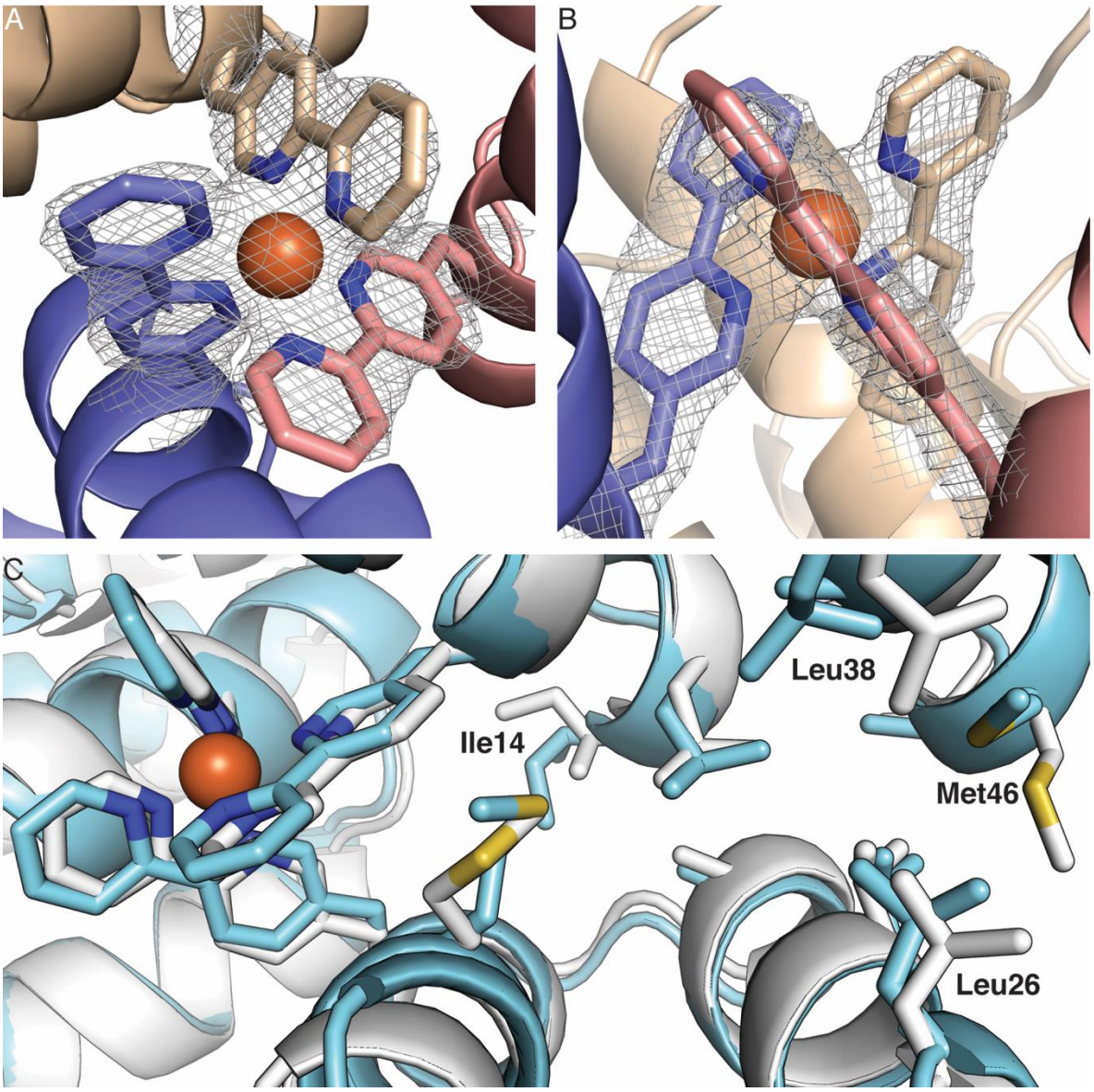

Fig. 3. X-ray crystallographic analysis of TRI_05. Electron density in the vicinity of the $[Fe(Bpy\text{-}ala)_3]^{2+}$ complex of TRI_05 is shown in (A) and (B) contoured to 1.5 σ. (C) An overlay of the design model (white) with the solved structure (blue) is shown in the vicinity of the designed interface. All designed residues in the interface are depicted as sticks. Residues whose side chains deviated from the designed conformation are labeled.

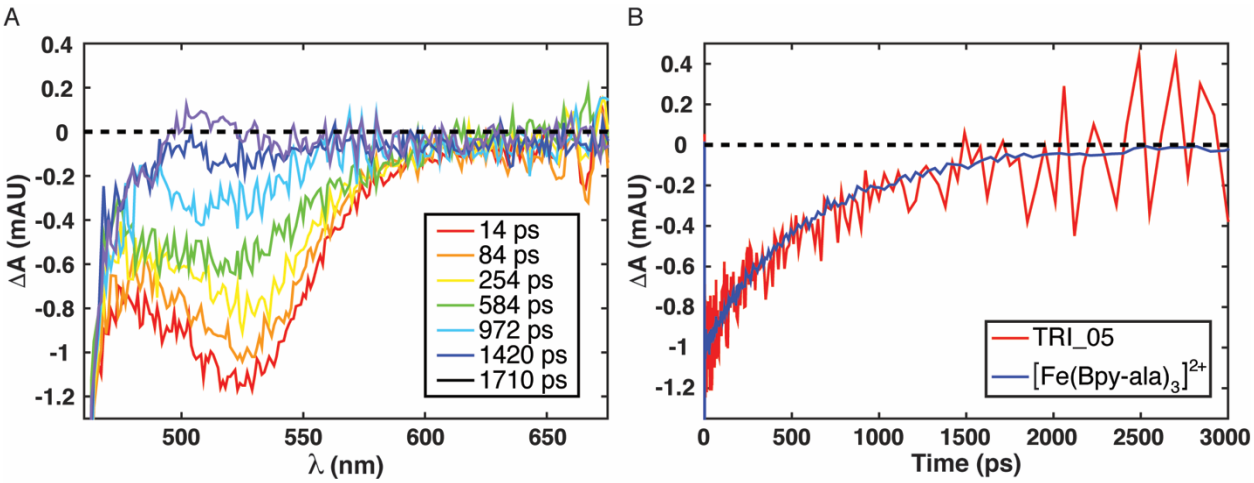

Fig. 4. Transient absorbance analysis of TRI_05. (A) Bleaching of TRI_05 MLCT absorbance at 7 distinct time delays after excitation at 440 nm. (B) Excited state lifetime of TRI_05 (red line) and $[Fe(Bpy\text{-}ala)_3]^{2+}$ (blue line) complexes after excitation at 440 nm. Absorbance measurements were obtained at 530 nm.